DOCUMENT RESUME

ED 237 505 SP 023 410

AUTHOR Kuhs, Therese; And Others
TITLE Differences among Teachers in Their Use of Curriculum-Embedded Tests. Research Series No. 133.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Aug 83
CONTRACT 400-81-0014
NOTE 29p.
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($3.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Behavioral Objectives; Case Studies; Educational Testing; *Elementary School Mathematics; *Elementary School Teachers; *Quantitative Tests; Student Evaluation; Teaching Methods; *Test Interpretation; Test Results; *Test Use
IDENTIFIERS *Curriculum Embedded Tests

ABSTRACT

Year-long case studies of instructional practices in elementary school mathematics were analyzed to describe testing practices of seven teachers in three school districts in Michigan and the uses those teachers made of test results. Focused on were: (1) curriculum-embedded tests (in particular the Management By Objectives system); (2) the use of results of curriculum-embedded tests for placement of students; (3) content decisions; and (4) student evaluation. Each time teachers made placement decisions (e.g., to create homogeneous groups within classrooms), some form of curriculum-embedded test was used. When teachers made content decisions about topic selection and pacing, four distinct styles of curriculum-embedded test usage emerged: (1) not monitoring student progress and referring to tests rarely, if ever; (2) relying on informal assessments (not labeled as tests) that are conducted with the intent of determining adequacy of student understanding; (3) using curriculum-embedded tests to make the domain for mastery clear and public but not actually to monitor student progress; and (4) using the results of curriculum-embedded tests as the sole criterion for deciding when a student can move on to a new topic. Overall, the data suggest that curriculum-embedded tests are important for some purposes, but not for others. (Author/JMK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237505

Research Series No. 133

DIFFERENCES AMONG TEACHERS IN THEIR USE OF CURRICULUM-EMBEDDED TESTS

Therese Kuhs, Andrew Porter, Robert Floden, Donald Freeman, William Schmidt, & John Schwille

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. C. Porter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

August 1983

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-81-0014)

SP 023 410

Institute for Research on Teaching

The **Institute for Research on Teaching** was founded at Michigan State University in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second contract to the IRT, extending work through 1984. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science, and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers, who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Co-Directors: Jere E. Brophy and Andrew C. Porter

Associate Directors: Judith E. Lanier and Richard S. Prawat

Editorial Staff
Editor: Janet Eaton
Assistant Editor: Patricia Nischan

Abstract

Year-long case studies of instructional practices in elementary school mathematics were analyzed to describe testing practices of teachers and the uses made of test results. The focus of the study was on curriculum-embedded tests; the use of results of curriculum-embedded tests for placement of students, content decisions, and student evaluation were given special attention. Every time teachers made placement decisions (e.g., to create homogeneous groups within classrooms), some form of curriculum-embedded test was used. In the making of content decisions about topic selection and pacing, four distinct styles of curriculum-embedded test usage emerged: (1) not monitoring student progress and referring to tests rarely, if ever; (2) relying on informal assessments (not labeled as tests) that are conducted with the intent of determining adequacy of student understanding; (3) using curriculum-embedded tests to make the domain for mastery clear and public but not to actually monitor student progress; and (4) using the results of curriculum-embedded tests as the sole criterion for deciding when a student can move on to a new topic. Overall, the data suggest that curriculum-embedded tests are important for some purposes, but not for others.

DIFFERENCES AMONG TEACHERS IN THEIR USE OF CURRICULUM-EMBEDDED TESTS

Therese Kuhs, Andrew Porter, Robert Floden,
Donald Freeman, William Schmidt, and John Schwille[1]

In spite of the continuing debates about testing, empirical studies of how teachers use tests are few. Most of these investigations of test usage consider only nationally available standardized tests (e.g., Kellegan, Madaus, & Airasian, 1982). However, the tests that are now most strongly recommended are not standardized tests, but tests that are clearly tied to the school curriculum and are an integral part of teachers' daily instruction (e.g., Tyler & White, 1979; Bloom, 1981; Edmonds, 1979). We use the term *curriculum-embedded* to describe such tests.

Perhaps curriculum-embedded tests have received so little attention in the research literature because they lack a clear definition. In the classroom it is often difficult to know what constitutes a test and what does not. Elementary school teachers can monitor student achievement more or less continuously. They observe students doing assignments and check the accuracy of completed assignments. One might conclude, therefore, that most teachers are testing some students most of the time. On the other hand, a textbook exercise labeled as a chapter test might be used by the teacher as just another exercise. Which is more the test, a class discussion that convinces the

---

[1]Theresa Kuhs, formerly a research intern with the IRT's Content Determinants Project, now teaches at the University of South Carolina. Andrew Porter, co-director of the IRT and associate dean of MSU's College of Education, coordinates the Content Determinants Project. Robert Floden, Donald Freeman, William Schmidt, and John Schwille are IRT senior researchers and members of the Content Determinants Project. Floden is an associate professor of teacher education; Freeman is a professor of teacher education; Schmidt is the chairperson of the Department of Counseling, Educational Psychology, and Special Education; and Schwille is a professor of teacher education.

teacher that students are confused about certain content or a chapter test given by a substitute teacher because the regular teacher doesn't trust the substitute to carry on normal instruction?

A second possible reason for researchers' lack of attention to curriculum-embedded tests may be that such tests are, in general, less controversial than externally required standardized tests. What little empirical research has been done on the uses of tests has generally been inspired by criticisms of externally required tests.

Whatever the reason for the lack of research on teachers' use of curriculum-embedded tests, these tests can and sometimes do represent an important part of the instructional process and are worthy of our attention. A review of teacher uses of assessment devices (Rudman, Kelly, Wanous, Mehrens, Clark, & Porter, 1980) found the only testing practices consistently tied to positive effects on student achievement were the numerous variations on the mastery learning theme (i.e., objectives are specified and student progress is consistently monitored through tests tied to those objectives). A recent National Institute of Education (NIE) conference on testing recommended more research and development on how to create instructional environments in which "testing is merged into the teaching process and provides timely and rich feedback to students, the teacher, and other appropriate parties" (Tyler & White, 1979, p. 5). Nationally available standardized tests cannot meet this need. These national tests can neither be tied to the instruction of an individual teacher nor given with the frequency the above statement implies.

A Center for the Study of Evaluation (CSE) study of teacher testing practices (Burry, Catterall, Choppin, & Dorr-Bremme, 1982) addresses teachers' use of curriculum-embedded tests in elementary school. This survey of fourth- and sixth-grade teachers in a national probability sample of school districts is

important to the understanding of current practices. For example, the CSE survey found teachers reporting an average of 23 tests per year when teaching elementary school mathematics (approximately one test every week and a half over a full school year). The teachers reported that nearly 80% of this testing involved curriculum-embedded tests.

Here, we take a much more microscopic view of curriculum-embedded tests and suggest that the CSE averages conceal much variation in teacher practice. Our data suggest that for some purposes, curriculum-embedded tests are important and for other purposes they are not.

Testing Within A Study of Teacher Content Decisions in Mathematics

Our data on teachers' use of curriculum embedded tests come from an intensive study of factors that may influence teachers' content decisions in elementary-school mathematics. We conducted year-long case studies of seven teachers in grades 3, 4, and 5. The teachers came from each of two schools in each of three Michigan school districts.

The three school districts were selected for differences in type and strength of district policies that could influence teacher content decisions in mathematics. Knoxport,[2] a large urban district, had a management by objectives (MBO) program in elementary school mathematics. Sawyer, a rural small-town district that had become somewhat suburban, had just adopted a new district-wide textbook series for use in elementary school mathematics. Finn, similar to Sawyer demographically, had a policy of building autonomy for mathematics content decision making. Within each district, we selected for study two schools that differed in the extent to which teachers worked in self-contained (and isolated) classrooms.

---

[2]Names of school districts, schools, and teachers have been changed.

The seven teachers who participated in the study were nominated by their building principals to meet the following specifications: (1) teach in grades 3 through 5, (2) be neither new to teaching (have at least three years of experience) nor near retirement, and (3) be willing to participate in the study.

In Sawyer, both teachers studied taught fourth grade. In Finn, we studied one third- and one fourth-grade teacher. In Knoxport, two fourth-grade teachers (in separate buildings) and one fifth-grade teacher were studied. In Knoxport, we selected two of these teachers from the same school to get a sense for possible within-school differences in content decision making.

Because this study focused on the mathematics instruction of just a few teachers and was not limited to a description of testing and test use, it has potential for making a unique contribution to the understanding of teachers' uses of curriculum-embedded tests. The case-study methodology provided data to support a thorough understanding of each teacher's approach to the teaching of mathematics. Since the data collected did not focus on testing more than on other aspects of instruction, data-collection procedures were unlikely to lead teachers to overreport test use or modify testing practice in response to the study.

Teachers kept daily logs of their instructional activities in mathematics for an entire school year (1979-80). In these logs they recorded the amount of time they spent on mathematics, their instructional goals, and the teaching strategies they used (including testing). They also noted all differences among students in content taught and provided copies of all materials used. Teacher logs were collected each week and checked for completeness.

The teachers were not always clear in their log reports as to whether or not they considered an instructional activity a test. The criteria used to

classify activities listed in the logs as tests were as follows: (a) the teacher referred to the activity as a test or (b) the activity made use of a test page from published instructional materials.

During weekly contacts, teachers were also interviewed about their use of textbooks, tests, objectives, and any conversations or newly-received materials related to mathematics. Lengthy interviews were conducted with the teachers at the beginning of the year to ascertain their intentions and priorities and at the end of the year to probe their reactions to possible curriculum influences. A limited amount of classroom observation was done to enhance our understanding of our other data.

While the focus on mathematics was useful in getting a full understanding of teachers' testing practices, it may be that some results will not apply to other subjects. For example, Yeh (1978) found that teachers test more frequently in mathematics than in reading. In her study of elementary school teachers from five California districts, approximately 78% of the teachers reported giving math tests once a month or more, while only 68% reported testing that frequently in reading. Grade level is another factor that potentially limits the generalizability of this study. Yeh found that there was an inverse relationship between grade level and amount of testing. Third- and fourth-grade teachers were the most frequent users of tests developed by others (e.g., textbook chapter tests), while fifth- and sixth-grade teachers were the most likely to develop their own tests. Still, 50% of the third- and fourth-grade teachers reported at least some test development.

The study is also limited by having been conducted in three districts in one area of Michigan. These districts do not represent the full range of variation in state and district test policies. Yeh (1978) leads us to suspect an inverse relationship between the frequency of use of curriculum-embedded

tests and the amount of district or state required testing. Moreover, in the same study, locally-developed (particularly teacher-made) tests were more often used for making instructional decisions other than placement, while required tests were used for reporting to others, particularly parents.

The state of Michigan requires testing in mathematics of all 4th, 7th, and 10th graders. At the time of our study, the state tests were objective-referenced and, for fourth-grade mathematics, contained 160 items to assess 33 objectives. The objectives were described by the state as minimal, and results were reported in terms of number of objectives mastered (where three of five items correct was taken as mastery).

In addition to required state testing, each of the three districts had its own testing program. In Knoxport, the Stanford Achievement battery was administered in all elementary grades each spring. Sawyer administered the Metropolitan Achievement Test each spring in third and sixth grade, while Finn used the Stanford to test students in grades 2 through 5 each fall.

Types of Test Use

In describing teachers' use of curriculum-embedded tests, we have tried to recognize when tests are not used as well as when they are. We have categorized the uses of tests as follows:[3]

1. for placement of students, including assignment to classrooms within a grade and assignment to instructional groups within a classroom;

2. for making decisions about what content topics are to be taught, for how long, and to what standards of performance; and

---

[3]The categories are taken, in part, from categories of test use identified by Tyler and White (1979), Yeh (1978), and, in part, our own conception of teacher content decisions (Schwille, Porter, Belli, Floden, Freeman, Knappen, Kuhs, & Schmidt, 1983).

3. for student evaluation, including grading and reporting to others.

Knoxport District

In Knoxport, curriculum-embedded tests were very important to Andy, Lucy, and Teri, the three teachers studied. To a great extent, the tests determined the content that students studied during the year. The Management By Objectives (MBO) system was the source of much of this curriculum-embedded testing in mathematics. The MBO system provided (1) tests that could be used at the beginning of the school year to place students at the appropriate objective in the hierarchy of objectives defined by the MBO program, (2) mastery tests for each objective, (3) review tests covering subsets of objectives, and (4) end-of-year tests for grades 4 through 6.

Placement of students. Each of the three Knoxport teachers used MBO curriculum-embedded tests to place all of their students in the district single-strand mathematics objectives. Their placement procedures were not the same, however, nor were the consequences.

The district required teachers to use the MBO placement test to identify children for Title I compensatory education programs. This test consisted of a few items related to each objective over the range of grade levels in the MBO system. Of the two teachers we studied in the same building, Lucy relied solely on this test in placing all her students in the MBO single-strand objectives, and Teri, in contrast, used placement-test results to select a different series of objective-referenced pretests for each student. Teri did not complete her elaborate placement procedure until nearly the end of the first semester.

Since the third teacher, Andy, did not teach in a Title I eligible school, he was not required to use the district placement test. Instead, in

placing his students, he relied on a district-provided locator test that covered fewer objectives but included a larger number of items per objective.

While the three teachers differed somewhat in how they made placement decisions, these decisions were similar in one important respect: Students within the same classroom differed greatly as to the objective on which they were to begin instruction. For example, one of Andy's students was placed on Objective 2, and, at the other extreme, two of his other students were placed on Objective 50.

In other respects, because the teachers differed in their instructional routines, their decisions about student placement had quite different consequences. In Andy's classroom, mathematics instruction was entirely self-paced. Once students had been placed in the MBO sequence of objectives, they worked independently on topics in the hierarchy until they mastered all the objectives up to a point determined by Andy (this point was always beyond grade level). At such time, students continued to study independently, pursuing an enrichment program that was not governed by the MBO system. Lucy and Teri, on the other hand, offered dual programs of mathematics instruction. In Lucy's classroom, students pursued self-paced studies based on MBO placement in the mornings. Then, later in the day, Lucy provided whole-class instruction using a different textbook from that used for work on the objectives. Originally, Teri planned to do whole-class instruction three days each week and to provide the MBO-guided, self-paced experience on the other two days. But, in part because of her more elaborate procedures for student placement, she did not begin individualized work on the MBO objectives until the end of February.

In Andy's classroom, the MBO placement was a major determinant of what each child studied during the year; hence, his use of the more thorough set

of tests for placement can be readily understood. In Lucy's classroom, the MBO placement affected only a portion of a child's experience with mathematics, and the mandated use of the less exacting MBO placement test was accepted. Ironically, Teri's demand for more accurate data postponed her placement decisions and made the consequences of placement less important.

Selection of topics and pacing. To the extent that teachers stayed within the MBO system, topic selection was controlled by the single-strand objectives, and pacing was monitored by performance on the objective-referenced posttests provided by the district. As has been noted, however, each of the three teachers used additional materials for instruction outside the MBO system.

Andy described what he called his enrichment program as the remainder of the fourth-grade textbook. Students worked on the enrichment program after finishing the prescribed MBO objectives for their grade level. Once students completed enrichment exercises, they returned to the MBO objectives for further work. Although individual students in his class were tested more often than students in any of the other classes studied, the testing was limited to MBO objectives.

In Andy's classroom, one-third of the students pursued the enrichment experiences in the text for over a quarter of the school year. During that time, no tests were administered. Students worked independently, and the only teacher-student interactions were student initiated. Thus, unless a student requested assistance, even informal assessments were not made.

During self-paced instruction in Lucy's classroom, decisions about topic selection and pacing were controlled by MBO objectives and tests. But Lucy also covered an entire, separate, fourth-grade textbook using whole-class instruction, presenting each lesson in the sequence in which it appeared in

the text. As did Andy in his enrichment program, Lucy chose not to test students on the material presented in these afternoon sessions.

Thus, Lucy's students were assessed only on MBO content. Further, because students were progressing at different speeds through the objectives, testing of a topic often did not coincide with the time the topic was presented in whole-class lessons. Clearly, any instructional decisions made concerning the afternoon sessions in Lucy's class were based on informal assessments of student progress--if, indeed, student mastery was considered at all. Because the entire textbook was covered, it seems unlikely that student mastery of topics was given much consideration in determining the pace of instruction for Lucy's afternoon sessions.

Like Lucy and Andy, Teri chose to limit formal assessments to MBO posttests. Unlike Lucy and Andy, she did not always tie the administration of a posttest to completion of instruction on the tested objective. Teri seemed more intent on identifying posttests students could pass than on identifying objectives for which students needed remedial instruction. For example, she told her students that anyone who mastered 25 objectives would be invited to her home for a pizza party at the end of the year. When the end of the school year neared and few students had reached this standard, Teri administered a large number of objective posttests to all her students.

The Knoxport teachers differed in how they decided when to give a student an objective-referenced posttest. At the beginning of the school year, Andy left it up to the students to decide when they were ready to take a mastery test. During the first week in February, Andy decided to exercise greater control over this decision. At first, he limited mastery testing to a specified day (each Friday). He explained that he was "seeing too much testing and failure and not enough willingness to do assignments." Two weeks

later, Andy took even greater control over when students could take a mastery test by requiring students to solve orally one or two problems before taking a posttest. Throughout the year, Lucy's strategy was similar to the final strategy Andy employed.

Teri, on the other hand, was more thorough in monitoring students' work on objectives. She collected and graded all completed objective-referenced assignments. Students were told at the beginning of each self-paced session whether they were to do another assignment or take a mastery test on the work they had done. Teri's decision to personally monitor students' success on assignments created a problem in that she could not review the assignments quickly enough for students to do self-paced work twice each week. As a result, only 26 lessons during the year were allocated to students' work on self-paced assignments, half of which occurred during the last month of school. For reasons described earlier, much of this time was spent taking tests. In fact, over these 26 lessons an average of 29 tests per student were taken.

Student evaluation and reporting to others. For the Knoxport teachers, the MBO tests were an important means of communicating what children had learned to persons outside the classroom. Student records of progress in the MBO system were reviewed by monitors sent out by the district's central administration and by the teachers the students would have next year. Further, all three teachers we studied reported on children's MBO profiles during parent conferences.

Sawyer District

In Sawyer, curriculum-embedded tests were not very important to the teachers studied (Jacqueline and Wilma), at least in comparison with their

importance to the three Knoxport teachers. The teachers' own convictions, the practices of other teachers, and, for one teacher, the new district textbook were much more important in influencing student opportunities to learn.

Placement of students. In Madison School, students were taught by teams of teachers. Jacqueline's fourth/fifth-grade team used ability grouping in mathematics, language arts, reading, and, to a limited extent, in science and social studies. Team members spent the first two weeks of school teaching mathematics to homeroom students, reviewing skills they believed to have been covered in the preceding year.

Decisions about placing students in instructional groups were made by team members after these two weeks of review. The team administered a placement test they had developed five years earlier that covered a wide range of topics, including place value; concepts and skills in addition, subtraction, and multiplication; story problem solving; and some division problems.

Jacqueline was assigned students who had done well on the placement test. For the most part, the membership of her mathematics group remained stable throughout the year. The primary exception to stable group membership was also an exception to the use of placement test results as the main criterion in creating homogeneous groups. Initially, one emotionally impaired and mathematically weak student was assigned to Jacqueline's mathematics class because she had earlier established good rapport with this student. Later in the year, two low-ability students were transferred to Jacqueline to make up a remedial group of three. During the entire year, only one student was moved from another group to Jacqueline's high-ability mathematics group because of better performance than placement test results had predicted. No students were moved out of her class to the lower ability group taught by one of the other members of the team.

For all but one month of the year, Jacqueline taught mathematics to two groups, one high-ability and one remedial. During that month, the high-ability students were split into two groups based on written assignments and Jacqueline's impression of which students were ready to go on to study long division and which needed further work on multiplication. Eventually, the two groups were brought back together because Jacqueline found it difficult to provide direct instruction to two high-ability groups plus the low-ability students.

Wilma, the other Sawyer teacher studied, was not a member of a teaching team and used only whole-class instruction. Thus, for Wilma, there were no placement decisions to make, with or without the aid of curriculum-embedded tests. However, placement decisions made in fifth grade, the next higher grade, did have some influence on her mathematics instruction. At the beginning of fifth grade, students were given a placement test written by fifth- and sixth-grade teachers in the building and used to homogeneously group students for the next two years of their mathematics instruction. Wilma deliberately emphasized content that she knew was on that placement test, explaining, "after all, when these children finish and go on to fifth grade next year, the first thing they are going to get hit with in the fall is a battery of tests measuring their skills in addition, subtraction, multiplication, and division."

Selection of topics and pacing. To determine whether or not teachers used tests to make pacing decisions, the teachers' log entries for days following activities we classified as tests were analyzed. For example, additional instruction on content just tested was taken as evidence that test performance led to remediation experiences.

Not counting tests of basic facts, Jacqueline recorded formal testing of her class 15 times during the year, or about once every two and a half weeks. Her log entries, however, provided little evidence of topic selection or pacing being based on those tests. Typically, on the day following a test, items were placed on the chalkboard as feedback to students, then new content was introduced. Only once was there substantial remediation after (as opposed to before) testing. On that occasion, Jacqueline focused review on a particular group of test problems. Students who had scored lower than *B* on the test were required to individually review incorrectly answered problems with the teacher.

For Jacqueline, chapter tests appeared to be an official certification of mastery (i.e., confirmation of what she already knew about student performance) rather than a diagnostic inquiry that might direct instruction. One evidence of this was Jacqueline's practice of occasionally postponing or even omitting testing for students she considered insufficiently prepared. For example, once she tutored three students while the class was taking a test. The three students then took the test later in the day. On another exceptional occasion, when her remedial group had been having great difficulty with the measurement chapter in the textbook, Jacqueline decided not to test at all, saying that the chapter was too hard for the students and for once she would have to be content with exposure rather than mastery.

In deciding whether students were ready to take tests and begin new topics, Jacqueline used day-to-day written assignments as well as students' general performance in class. In several ways these assignments were, for Jacqueline, formal assessments. She kept careful records, and students had to complete assignments, whether they were absent or simply slow. Student performance on these assignments influenced pacing. Ordinarily, in the year of

our study, the high-ability students in Jacqueline's class moved along at a pace she considered satisfactory. Students who had occasional difficulty received remediation; students who finished assignments early often did extra-credit assignments. Jacqueline routinely recorded percent grades on assignments for communication to students and use in computing report-card averages.

The importance of these written assignments is illustrated in Jacqueline's decision to split her high-ability group for about a month in the late winter, letting some students go on to a new topic while keeping others on two-digit multiplication. For some time she had been uncertain about how to deal with the varied progress of her high-ability students on two-digit multiplication, the first snag she reported having with this group all year. At first she hoped to avoid splitting the group by continuing to work individually with the students who were having difficulty. She thought a particular ditto (Holt ditto 183) would make clear whether students were ready to go on. But the decision was delayed still further; so many children were out sick due to a flu epidemic that Jacqueline said she could not tell how they were doing until she got their make-up papers. Yet the formal chapter test did not influence her decision; Jacqueline did not administer the multiplication chapter test to any of these students until a few days after she had split her high-ability group.

Unlike Jacqueline, Wilma did not follow the textbook closely. For example, Wilma began the year by covering the first 10 pages of the textbook, but then skipped to page 59. From there, she skipped to page 50, then to page 13, next to page 57, then back to page 20, and so on through the year. Given this style of textbook usage, it was not surprising that she made no use of textbook chapter tests.

Wilma did report 11 situations in her log that were classified as tests. Consistent with her emphasis on computational skills, six of the 11 curriculum-embedded tests were on basic number facts (e.g., multiplication tables).

Other lessons in Wilma's classroom, however, did seem to represent occasions where student progress was assessed. This inference is based on annotations Wilma made in her log entries about activities she did not designate as involving a test. The following are examples of this:

January 30	Assignment to see how well students could divide without help.
April 7	Review--"to see what students remembered from before vacation."
May 13	Started lesson with independent assignment "to see how much students remembered."
May 22	Assignment "to see how much students remembered from concepts taught throughout the year."

Nevertheless, Wilma's testing practices and less-formal assessments of student progress did not change the pacing decisions she made at the beginning of the year. In an early interview she listed mathematics content areas and the dates she projected for beginning instruction. She called this her "subject clock," and it was based on five years of teaching experience, four of which were at her current school. Table 1 reports Wilma's projected starting dates and actual starting dates taken from the logs. In all cases, the projected and actual dates were within one to two weeks of each other.

Table 1

Accuracy of Wilma's Subject Clock

Content Area	Projected Starting Date	Actual Date from Log
Addition (and Place Value)	August 28	August 28
Subtraction	October 15	October 9
Multiplication	November 15	November 5
Division	February 1	January 21
Fractions & Other Peripheral Topics	April 7	April 16

Student evaluation and reporting to others. Of the two teachers in Sawyer, only Jacqueline made explicit use of curriculum-embedded tests for grading and to communicate learning outcomes in mathematics to persons outside the classroom. In October Jacqueline explained to her students that 25% of their report-card grade for mathematics would be based on test performance, with the other 75% being based on average percent grades in daily assignments.

Jacqueline's clearest use of a test to communicate student performance to parents occurred in a unit on telling time. When another member of the team identified telling time as an area of difficulty for students, the team jointly planned four sessions of supplementary math instruction in which the teachers' aides supervised instruction on telling time. Parents of students who did poorly on a posttest developed by yet another member of the team were sent letters indicating their child's continuing deficiency and asking that they give their child practice at home on telling time.

Finn District

In Finn, as in Sawyer, the teachers (Donna and Bob) made limited use of curriculum-embedded tests. Donna, who taught in an open-space school, used a

test to place students in small ability groups but administered tests infrequently after the groups were formed. Bob, who taught a self-contained classroom in Roosevelt School, made some use of tests for pacing but questioned the validity of certain chapter tests that he might have used.

Placement of students. No attempt was made in either of the two schools to create homogeneous ability groups through initial assignment of students to classrooms. Student placement was never an issue in Bob's class because, for the most part, he taught new topics to the whole class.

Donna used whole-class instruction until November, when she divided her class into two instructional groups. Prior to the formation of these groups, students completed four timed tests on multiplication tables and two other tests covering a broader range of content. Donna also had the results from the fall Stanford Achievement testing. Nevertheless, she assigned students to instructional groups for mathematics based solely on a timed test of 100 multiplication facts (i.e., a test of 100 items of the type 8 x 9 = ?).

Student performance on this curriculum-embedded test had substantial consequences. The higher ability group received instruction on topics that were not presented to the lower ability group (e.g., fractions, two-digit division, and problems with three-digit multipliers). Further, Donna's comments reflected that she had different expectations for the lower ability group. For example, she explained that students in the lower ability group could not think about too many things at any one time.

Selection of topics and pacing. Both Bob and Donna focused their testing on the content they viewed as most important for students to master (e.g., basic number facts). Nine of the 12 basic-number-fact tests Bob used were mastery quizzes from individualized packets of materials he had assembled.

Further, both teachers required each student to demonstrate mastery of multiplication tables in an oral quiz. Interviews with these teachers revealed that they viewed memorization of basic facts as one of the more important outcomes of mathematics studies.

Bob's interpretation of several test results further demonstrated that his view of the relative importance of topics affected the way he used test results. Because he followed the scope and sequence of the textbook closely, it was convenient for him to use the chapter tests from the book for assessment. However, Bob frequently discounted the importance of student errors on these tests. For example, a chapter test on subtraction included problems that were written as missing-addend problems. While he had gone over the textbook's treatment of the topic, Bob explained he was not interested in knowing whether or not students could do missing-addend problems. In another instance, Bob said his students "really bombed" on a story-problem chapter test. Bob reported that the students had difficulty because the test used the phrase "how many does this make" instead of "how many in all." Since Bob felt it pedagogically important not to confuse students with several different problem formats, he viewed students' difficulty as a problem with the book rather than as student failure.

In addition to the focus on computation, which she viewed as important, Donna's log entries revealed a particular pattern of testing. Before instructional groups were formed in November, Donna reported nine test events (six of these were tests of basic multiplication facts). After the class was divided into groups, testing was less frequent (i.e., only three tests in the high-ability group and two in the low-ability group). Further, two of the three tests administered to the high-ability group were given on sequential days

when a substitute teacher was in the class. Instruction on days following tests did not deal extensively with the content tested.

Despite her limited use of tests, Donna explained that she knew when students were ready to go on to something new by observing the accuracy of their work and by their ability to switch from one problem type to another without errors. In her log, Donna reported many assessments that were not labeled as tests. Below are some examples:

December 6	"Students are having a rough time doing the shorter method."
January 4	"Group is still forgetting how to work multiplication problems ex. 36 x 48 and need more work at this."
January 22	"The class caught on to division very quickly. It only seemed necessary to do one type of problem on page 194."
February 12	"The group is still working on math packet mainly because of errors students are making in multiplying. They can do the process."

Student evaluation and reporting to others. The only evidence that Donna or Bob used test results to report student progress to others occurred early in the school year. In October, Donna commented that the teacher-made test she gave "was mainly for report card evaluation."

Since Bob frequently discounted the results of the tests he gave, he did not consider a total test score valid for use in reporting to others. He did, however, discuss students' performance on the Stanford Battery with parents during parent conferences, and he indicated that the parents showed considerable interest in those results.

Summary and Discussion of Findings

In every case where placement decisions were made by teachers in the present study, either to homogeneously group students into classrooms or to

form homogeneous groups within classrooms, some form of curriculum-embedded test was used. Of the three a priori categories of test usage that formed the framework for analysis, student placement provided the most frequent examples of teacher reliance on curriculum-embedded test results.

The three Knoxport teachers used district MBO tests to determine students' starting place for self-paced studies. Teri used an even more elaborate system of testing than the MBO system prescribed, and Andy chose to use the more detailed of the two available placement tests. The teachers who chose to identify homogeneous instructional groups (Jacqueline, who grouped students across the teamed classrooms, and Donna, who subdivided her class into two groups) based student group membership on the results of a formal test. However, both teachers conducted instruction without grouping students for a period of time before administering the placement test, allowing informal observation to influence their placement decisions. Although the other two teachers did not make placement decisions, one of them was in a school where teachers at the next higher grade relied on a locally developed test for grouping students in mathematics.

When making decisions about what topics to teach and how quickly to move from one topic to the next, four distinct styles of test usage emerged. Teachers displaying the first style did not monitor student progress and rarely, if ever, referred to tests. Each of the three Knoxport teachers, Lucy, Andy, and Teri, used this style when providing instruction not included in the district's MBO system. Teachers displaying the second style relied primarily on informal assessment, which they did not label as testing, to determine adequacy of student understanding. Wilma, Donna, and Bob were most clearly identifiable with this style. The teacher displaying the third style used tests to make the domain for mastery both clear and public,

but test results were redundant as far as making decisions about pacing and topic selection were concerned. For example, Jacqueline used chapter tests to mark the end of instruction on a topic and to confirm her own day-by-day assessment of student progress. We observed the fourth style of test usage in the self-paced MBO components of Andy and Lucy's mathematics instruction. There, tests were the sole criterion for determining when a student could move on to the next instructional objective.

Because the amount of curriculum-embedded testing varied considerably among the seven teachers, it is possible to speculate on the reasons teachers do or do not use tests. One plausible and parsimonious explanation is convenience. The Knoxport teachers were expected to provide instruction on the MBO mathematics objectives and were provided with copies of tests to use that were tied to those objectives. These teachers did more testing than did the other teachers, even when the testing was not required.

Nevertheless, these same teachers did almost no testing when providing instruction outside the MBO system where tests were less convenient. In addition, teachers who followed a textbook closely were more inclined to use chapter tests than were other teachers.

There were, however, a few instances in which teachers constructed their own tests to serve a particular function. As already noted, placing students in homogeneous groups for instruction was one occasion. Teachers also saw timed tests as serving a useful function in getting students to memorize basic-skill facts. Finally, there was some evidence that the teachers were more inclined to construct tests for use when making decisions about students that affect other teachers (i.e., shared decision making) as in the case of Jacqueline's team.

There was little evidence in our case studies that the teachers used curriculum-embedded tests to make grading decisions. The CSE survey, however, found 95% of the elementary teachers reporting that they used self-made tests and 77% reporting they used chapter tests for purposes of determining student grades in mathematics (Burry et al., 1982). Except for Jacqueline, the teachers we studied appeared to make grading decisions based on general perceptions of daily student performance. Tests are a part of daily performance, and perhaps this overlap explains the high rate of test usage in the CSE results. Nevertheless, the results of our case studies call into question the notion that tests play an important role in determining grades in mathematics.

In contrast to curriculum-embedded tests, externally-required tests given once a year or less were rarely used by teachers for any of the three categories of test use considered (Schmidt, 1981). The only placement decisions based on externally-required tests were for identifying students to participate in pull-out programs (e.g., gifted and compensatory education), and these decisions were not the teachers' to make. As for pacing, the frequency of the externally-required tests was insufficient for this purpose, and the teachers' topic selection did not seem to be much influenced by the state minimum competencies test or the district-used standardized tests (Schmidt, 1981). In fact, surprisingly few teachers had a thorough knowledge of the mathematics content on standardized tests that had been given for some time in their district. Some of the teachers did provide feedback to parents on student performance on state and standardized tests. However, because the tests were not viewed as tied to instruction, test performance was not taken as an indicator of performance in a particular teacher's class.

Conclusion

While the case studies revealed relatively little testing and use of test results when making instructional decisions among teachers, neither was there evidence that teachers would resist attempts to increase the amount and use of testing in their classrooms. Over 70% of the teachers in the CSE survey reported beliefs that testing motivates students to study harder and that more competency, proficiency, and functional-literacy testing should be done (Burry et al., 1982).

Those who wish to increase the frequency of formal testing in classrooms should consider the situations that, in this study, encouraged teachers to test. The use of an objective-referenced instructional management system resulted in more frequent testing. However, the Knoxport MBO system did not cover the entire range of content taught, and content outside the MBO system was not tested. Further, because a self-paced instructional strategy was used, some students were tested less often than students in classrooms in other districts (e.g., one student only completed three posttests during the year). Teachers who followed a textbook closely also tested regularly, but they did not use test results to a great extent to inform their instructional decisions. If classroom testing practices are to be changed, teachers must be made more aware of the benefits and pitfalls of frequent testing and learn more about how to design and use a testing program in their classrooms.

References

Bloom, B. S. (1981). All our children learning: A primer for parents, teachers and other educators. New York: McGraw-Hill Book Co.

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. (1982). Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs? (CSE Report No. 194). Los Angeles: Center for the Study of Evaluation, University of California.

Edmonds, R. (1979). Some schools work and more can. Social Policy, 9(5), 28-32.

Kelleghan, T., Madaus, G., & Airasian, P. (1982). The effects of standardized testing. Boston: Kluwer-Nijhoff Publishing.

Rudman, H., Kelly, J., Wanous, D., Mehrens, W., Clark, C., & Porter, A. (1980). Integrating assessment with instruction: A review (1922-1980) (Research Series No. 75). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Schmidt, W. H. (1981). Teacher knowledge and use of standardized tests. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., Knappen, L., Kuhs, T., & Schmidt, W. (1983). Teachers as policy brokers in the content of elementary school mathematics. In L. Shulman & G. Sykes (Eds.), Handbook of teaching and policy. New York: Longman, Inc.

Tyler, R. W., & White, S. H. (1979). Testing, teaching and learning: Report of a conference on research on testing. Washington, DC: National Institute of Education, U.S. Department of Health, Education, and Welfare.

Yeh, J. P. (1978). Test use in schools. Washington, DC: National Institute of Education, U.S. Department of Health, Education and Welfare.